In its interpretation of the terms of reference, the PPR Panel will then develop a Scientific Opinion to address the methodological limitations identified in epidemiological studies on pesticides and to make recommendations to the sponsors of such studies on how to improve them in order to facilitate their use for regulatory pesticide risk assessment, particularly for substances in the post-approval period.

This Scientific Opinion is intended to assist the peer review process during the renewal of pesticides under Regulation 1107/2009 where the evaluation of epidemiological studies, along with clinical cases and poisoning incidents following any kind of human exposure, if available, represent a data requirement. Epidemiological data concerning exposures to pesticides in Europe will not be available before first approval of an active substance (with the exception of incidents produced during the manufacturing process, which are expected to be very unlikely) and so will not be expected to contribute to a DAR. However there is the possibility that earlier prior approval has been granted for use of an active substance in another jurisdiction and epidemiological data from that area may be considered relevant. Regulation (EC) No 1107/2009 requires a search of the scientific peer-reviewed open literature, where it is expected to retrieve existing epidemiological studies. It is therefore recognised that epidemiological studies are more suitable for the renewal process of active substances, also in compliance with the provision of the EC regulation 1141/2010 indicating that "The dossiers submitted for renewal should include new data relevant to the active substance and new risk assessments to reflect any changes in data requirements and any changes in scientific or technical knowledge since the active substance was first included in Annex I to Directive 91/414/EEC".

The PPR Panel will specifically address the following topics:

1. Review inherent weaknesses affecting the quality of epidemiological studies (including gaps and limitations of the available pesticide epidemiological studies) and their relevance in the context of regulatory pesticide risk assessment. How can these weaknesses be addressed?
2. What are potential contributions of epidemiological studies that complement classical toxicological studies conducted in laboratory animal species in the area of pesticide risk assessment?
3. Discuss and propose a methodological approach specific for pesticide active substances on how to make appropriate use of epidemiological studies, focusing on how to improve the gaps and limitations identified.
4. Propose refinements to practice and recommendations for better use of the available epidemiological evidence for risk assessment purposes. Discuss and propose a methodology for the integration of epidemiological information with data from experimental toxicology.

1.4. Additional information

In order to fully address topics 1-4 above (section 1.3) attention has been paid to a number of relevant reviews of epidemiological studies and the experience of other National and International bodies with knowledge of epidemiology in general and in applying epidemiology to pesticide risk assessment specifically. Detailed attention has been given to these studies in Annex A and drawn from the experience of the authors that have contributed constructively to understanding in this area. Also Annex A records published information that has been criticised for its lack of rigour showing how unhelpful some published studies may be. The lessons learned from such good (and less-good) practice have been incorporated into the main text by cross-referring to Annex A. In this way this Scientific Opinion has the aim of clearly distilling and effectively communicating the arguments in the main text without overwhelming the reader with all the supporting data which is nevertheless accessible.

In addition, Annex B contains a summary of the main findings of a project that EFSA outsourced in 2015 to further investigate the role of human biological monitoring (HBM) in occupational health and safety strategies as a tool for refined exposure assessment in epidemiological studies and to contribute to the evaluation of potential health risks from occupational exposure to pesticides.

2. General framework of epidemiological studies on pesticides

This chapter introduces the basic elements of epidemiological studies on pesticides and contrasts them with other types of studies.

2.1. Study design

Epidemiology studies the distribution and determinants of diseases in human or other target species populations, to ascertain how, when and where diseases occur. This can be done through observational studies and intervention studies (i.e., clinical trials) [8]. Both types of studies are carried out in a natural setting, which is a less controlled environment than laboratories. To identify disease determinants that are associated with either the presence of disease (prevalence) or with the occurrence of new cases of disease over time (incidence). This is done by comparing study groups subject to differing exposure to a potential risk factor.

Information on cases of disease occurring in a natural setting can also be systematically recorded in the form of case reports or case series of exposed individuals only. Although case series/reports do not compare study groups according to differing exposure they may provide useful information, particularly on acute effects following high exposures, which makes them potentially relevant for risk assessment.

In clinical trials the exposure of interest is randomly allocated to subjects and, whenever possible, these subjects are blinded to their treatment, thereby eliminating potential bias due to their knowledge about their exposure to a particular treatment. This is why they are called intervention studies. Observational epidemiological studies differ from clinical studies in that the exposure of interest is not randomly assigned to the subjects enrolled and participants are often not blinded to their exposure. This is why they are called observational. As a result, randomized clinical trials rank higher in terms of design as they provide unbiased estimates of average treatment effects.

The lack of random assignment of exposure in observational studies represents a key challenge, as other risk factors that are associated with the occurrence of disease may be unevenly distributed between those exposed and non-exposed. This means that known confounders need to be measured and accounted for. However, there is always the possibility that unknown confounders are left unaccounted for (automatically accounted for in randomised clinical trials by their design). Furthermore the fact that study participants are often aware of their current or past exposure or may not recall these accurately in observational studies (e.g. second-hand smoke, dietary intake or occupational hazards) may result in biased estimates of exposure if it is based on self-report. As an example it is not unlikely that when cancer cases and controls are asked whether they have previously been exposed to a pesticide the cancer cases may report their exposure differently from controls, even in cases where the past exposures did not differ between the two groups.

Traditionally, designs of observational epidemiological studies are classified as either ecological, cross-sectional, case-control or cohort studies. This approach is based on the quality of exposure assessment and the ability to assess directionality from exposure to outcome. These differences largely determine the quality of the study (Pearce 2012; Rothman and Greenland 1998).

Ecological studies are observational studies where either exposure, outcome or both are measured on a group but not at individual level and the correlation between the two is then examined. Most often, exposure is measured on a group level while the use of health registries often allows for extraction of health outcomes on an individual level (cancer, mortality). These studies are often used when direct exposure assessment is difficult to achieve and in cases where large contrast in exposures are needed (comparing levels between different countries or occupations). Given the lack of exposure and/or outcome on an

[8] In this opinion, "human data" includes observational studies, also called epidemiological studies, where the researcher is observing natural relationships between factors and health outcomes without acting upon study participants. Vigilance data also fall under this concept. In contrast, interventional studies are outside the scope of this Opinion. These studies also called experimental studies or randomized clinical trials, and their main feature is that the researcher intercedes as part of the study design.

individual level, these studies are useful for hypothesis generation but results generally need to be followed up using more rigorous design in either humans or use of experimental animals.

In **cross-sectional studies** exposure and health status are assessed at the same time, and prevalence rates (or incidence over a limited recent time) in groups varying in exposure are compared. In such studies, the temporal relationship between exposure and disease cannot be established since the current exposure may not be the relevant time window that leads to development of the disease. Cross-sectional studies may nevertheless be useful for risk assessment if exposure and effect occur more or less simultaneously or if exposure does not change over time.

Case-control studies examine the association between estimates of past exposures among individuals that already have been diagnosed with the outcome of interest (e.g., cases) to a control group of undiagnosed subjects from the same population. In population-based incident case-control studies, cases are obtained from a well-defined population, with matched controls selected from members of the population who are disease free at the time a case is incident. The advantages of case-control studies are that they require less sample sizes, time and resources compared to prospective studies when studying rare outcomes such as some types of cancer. In case-control studies past exposure is most often not assessed based on 'direct' measurement but rather through less certain measurements such as a recall captured through interviewer or self-administered questionnaires or proxies such as job descriptions titles or task histories. Besides the main limitation that case control studies are prone to is recall-bias when estimating exposure, other challenges include the selection of appropriate controls; as well as the need for appropriate confounder control.

- In **cohort studies** the population under investigation consists of individuals who are at risk of developing a specific disease or health outcome at some point in the future. At baseline and at later follow-ups (prospective cohort studies) relevant exposures, confounding factors and health outcomes are assessed. After an appropriate follow-up period the frequency of occurrence of the disease is compared among those differently exposed to the previously assessed risk factor of interest. Cohort studies are therefore by design prospective as the assessment of exposure to the risk factor and covariates of interest are measured before the health outcome has occurred. Thus they can provide better evidence for causal associations compared to the other designs mentioned above. In some cases, cohort studies may be based on estimates of past exposure. Such retrospective exposure assessment is less precise than direct measure and prone to recall-bias. As a result the quality of evidence from cohort studies varies according to the actual method used to assess exposure and the level of detail by which information on covariates were collected. Cohort studies are particularly useful for the study of relatively common outcomes. If sufficiently powered in terms of size, they can also be used to appropriately address relatively rare exposures and health outcomes. Prospective cohort studies are also essential to study different critical exposure windows. An example of this is longitudinal birth cohorts that follow children at regular intervals until adult age. Cohort studies may require a long observation period when outcomes have a long latency prior to onset of disease. Thus, such studies are both complex and expensive to conduct and are prone to loss of follow-up.

2.2. Population and sample size

A key strength of epidemiological studies is that they study diseases in the very population about which conclusions are to be drawn, rather than a proxy species. However, only rarely will it be possible to study the whole population. Instead a sample will be drawn from the reference population for the purpose of the study. As a result the observed effect size in the study population may differ from that in the population if the former does not accurately reflect the latter. However, observations made in a non-representative sample may still be valid within that sample but care should then be made when extrapolating findings to the general population. Representative samples can be achieved through use of appropriate sampling schemes.

Having decided how to select individuals for the study, it is also necessary to decide how many participants should minimally be enrolled. The sample size of a study should be large enough to warrant sufficient statistical power (e.g. 80%). This is the likelihood that an effect of a magnitude that is considered biologically relevant or relevant from a regulatory perspective will also be statistically significant. For example, a power of 80% means that the study will confirm a true association with a probability of 80%. Also, small samples are likely to constitute an unrepresentative sample The statistical power is also closely related to risk inflation, which needs to be given special attention when interpreting results from small or underpowered studies (see Annex D).

Epidemiological studies, like toxicological studies in laboratory animals, are often designed to examine multiple endpoints unlike clinical trials that are designed and conducted to test one single hypothesis, e.g. efficacy of a medical treatment. To put this in context, for laboratory animal toxicology test protocols, OECD guidance for pesticides may prescribe a minimum number of animals to be enrolled in each treatment group. This does not guarantee adequate power for any of the multitude of other endpoints being tested in the same study. It is thus important to ascertain the power of a study post-hoc both in epidemiology and laboratory studies.

2.3. Exposure

The quality of the exposure measurements influences the ability of a study to correctly ascertain the causal relationship between the (dose of) exposure and a given adverse health outcome.

In toxicological studies in laboratory animals the 'treatment regime' i.e. dose, frequency, duration and route are well defined beforehand and its implementation can be verified. This often allows expression of exposure in terms of external dose administered daily via oral route for example in a 90-day study, by multiplying the amount of feed ingested every day by a study animal with the intended (and verified) concentration of the chemical present in the feed. Also, in the future, the internal exposure has to be determined in the pivotal studies.

In the case of pesticides, estimating exposure in a human observational setting is difficult as the dose, its frequency and duration over time and the route of exposure are not controlled and not even well known.

Measuring the intensity, frequency and duration of exposure is often necessary for investigating meaningful associations. Exposure may involve a high dose over a relatively short period of time, or a low-level prolonged dose over a period from weeks to years. While the effects of acute, high-dose pesticide exposure may appear within hours or days, the effects of chronic, low-dose exposures may not appear until years later. Also a disease may require a minimal level of exposure but increase in probability with longer exposure.

There may be differences in absorption and metabolism via different routes (dermal, inhalation and oral). While dermal or inhalation are often the routes exposure occurs in occupational settings, ingestion (food, water) may be the major route of pesticide exposure for the general population.

2.4. Health outcomes

The term health outcome refers to a disease state, event, behaviour or condition associated with health that is under investigation. Health outcomes are those clinical events (usually represented as diagnosis codes, i.e. International Classification of Diseases ICD-10) or outcomes (i.e., death) that are the focus of the research. Use of health outcomes requires a well-defined case definition, a system to report and record the cases and a measure to express the frequency of these events.

A well-defined case definition is necessary to ensure that cases are consistently diagnosed, regardless of where, when and by whom they were identified and thus avoid misclassification. A case definition involves a standard set of criteria, which can be a combination of clinical symptoms/signs, which can be supplemented by confirmatory diagnostic tests with their known sensitivity and specificity. The sensitivity of the whole testing procedure (i.e. the probability that a person with an adverse health condition is truly diagnosed) must be known to estimate the true prevalence or incidence.

The clinical criteria often involve a combination of symptoms and possibly other characteristics (e.g. age, occupation) that are associated with increased disease risk. At the same time, appropriately measured and defined phenotypes or hard clinical outcomes add validity to the results.

Mortality, cancer and other nation-wide health registries generally meet the case-definition requirements and provide (almost) exhaustive data on the incident cases within a population. These health outcomes are recorded and classified in national health statistics databases, which depend on accepted diagnostic criteria that are evolving and differ from one authority to another. Also, diagnoses can be recorded in refined or relatively crude format. This may confound attempts to pool data usefully for social benefit.

Although the disease status is typically expressed as a dichotomous variable, it may also be measured as an ordinal variable (e.g., severe, moderate, mild or no disease) or as a quantitative variable for example by measuring molecular biomarkers of toxic response in target organs or physiological measures such as blood pressure or serum concentration of lipids or specific proteins.

The completeness of the data capture and its consistency are key contributors to the reliability of the study. Harmonisation of diagnostic criteria, data storage and utility would bring benefits to the quality of epidemiological studies.

A surrogate endpoint is used as substitute for a well-defined disease endpoint, an outcome measure, commonly a laboratory measurement (biomarker of response). These measures are considered to be on the causal pathway for the clinical outcome. In contrast to overt clinical disease, such biological markers of health may allow to detect subtle, subclinical toxicodynamic processes. For such outcomes, detailed analytical protocols for quantification should be specified to enable comparison or replication across laboratories. The use of adverse outcome pathways can highlight differences in case definitions (EFSA 2017).

Although surrogate outcomes may offer additional information, the suitability of the surrogate outcome examined needs to be carefully assessed. In particular, the validity of surrogate outcomes may represent a major limitation to their use (la Cour et al., 2010). Surrogate endpoints that have not been validated should thus be avoided.

When the health status is captured in other ways, such as from self-completed questionnaires or telephone interviews, from local records (medical or administrative databases) or through clinical examination only ,these should be validated to demonstrate that they reflect the underlying case definition.

2.5. Statistical analysis and reporting

Reporting in detail materials, methods and results, and conducting appropriate statistical analyses are key steps to ensure quality of epidemiological studies. Regarding statistical analysis, one can distinguish between descriptive statistics and modelling of exposure-health relationships.

2.5.1. Descriptive statistics

Descriptive statistics aim to summarize the important characteristics of the study groups, such as exposure measures, health outcomes, possible confounding factors and other relevant factors. The descriptive statistics often include frequency tables and measures of central tendency (e.g. means and medians) and dispersion (e.g. variance and interquartile range) of the parameters or variables studied.

2.5.2. Modelling exposure-health relationship

Modelling of the exposure-health relationship aims to assess the possible relationship between the exposure and the health outcome under consideration. In particular, it can evaluate how this relationship may depend on dose and mode of exposure and other possible intervening factors.